PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-222540
(43)Date of publication of application : 21.08.1998

(51)Int.Cl. G06F 17/30

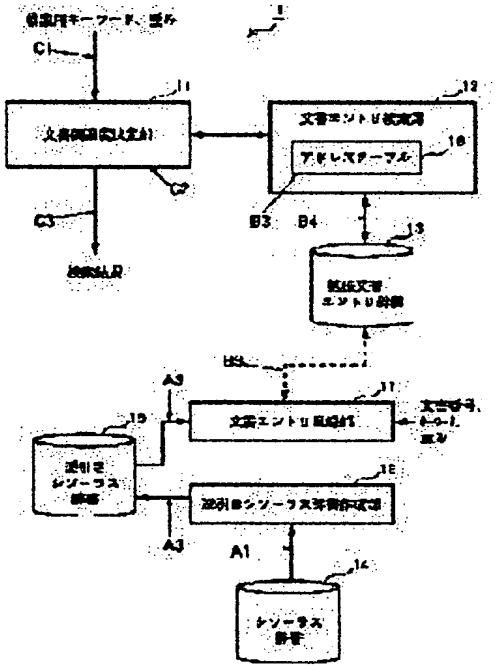
(21)Application number : 09-330453
(22)Date of filing : 01.12.1997

(71)Applicant : N T T DATA TSUSHIN KK
(72)Inventor : NAKAJIMA HIROYUKI
KITANI TSUYOSHI

(30)Priority
Priority number : 08323731 Priority date : 04.12.1996 Priority country : JP

(54) DOCUMENT RETRIEVING METHOD, DEVICE AND RECORDING MEDIUM

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a document retrieving device which improves the processing speed of document retrieval.
SOLUTION: An extended document entry dictionary 13 stores the document number and weight of a related document in each keyword. The correspondence of a keyword and a document number is preliminarily undergone thesaurus expansion. When a document association degree deciding part 11 inputs a set of a retrieving keyword and weight, a document entry retrieving part 12 reads the document number and weight that correspond to an inputted retrieving keyword from the dictionary 13. The part 11 calculates the degree of association of a retrieving keyword and a document in each read document number and outputs a document number in order of the degree of association.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号

特開平10－222540

(43)公開日　平成10年(1998)８月21日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｇ０６Ｆ　17/30		Ｇ０６Ｆ　15/40	３７０Ａ
		15/403	３４０Ａ
			３５０Ｃ

審査請求　未請求　請求項の数８　ＯＬ　（全 11 頁）

(21)出願番号　特願平9－330453

(22)出願日　平成９年(1997)12月１日

(31)優先権主張番号　特願平8－323731

(32)優先日　平８(1996)12月４日

(33)優先権主張国　日本（ＪＰ）

(71)出願人　000102728
エヌ・ティ・ティ・データ通信株式会社
東京都江東区豊洲三丁目３番３号

(72)発明者　中島　浩之
東京都江東区豊洲三丁目３番３号　エヌ・ティ・ティ・データ通信株式会社内

(72)発明者　木谷　強
東京都江東区豊洲三丁目３番３号　エヌ・ティ・ティ・データ通信株式会社内

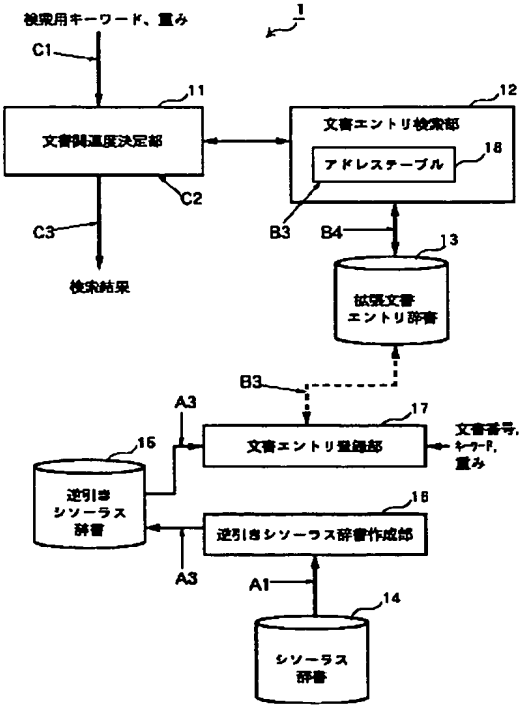
(74)代理人　弁理士　鈴木　正剛

(54)【発明の名称】　文書検索方法、装置及び記録媒体

(57)【要約】

【課題】　文書検索の処理速度を向上させる文書検索装置を提供する。

【解決手段】　拡張文書エントリ辞書１３に、キーワード毎に関連する文書の文書番号と文書の重みとを格納しておく。キーワードと文書番号との対応は、予めシソーラス展開されたものである。文書関連度決定部１１が検索用キーワードと重みの組を入力すると、文書エントリ検索部１２は、拡張文書エントリ辞書１３から入力された検索用キーワードに対応する文書番号と文書の重み読み出す。文書関連度決定部１１は、読み出された文書番号毎に、検索用キーワードと文書との関連度を算出し、関連度の高い順に文書番号を出力する。



【特許請求の範囲】

【請求項１】　コンピュータを用いた文書検索方法であって、検索対象となる文書の識別情報、当該文書に含まれるキーワードと検索目的に応じて当該キーワードに付与される重み、及び、前記キーワードが関連する他のキーワードとその関連度を取得して各キーワードに対する文書の重みを算出し、算出された重みを当該文書の識別情報と共にキーワード毎に設定されたメモリ領域へ蓄積する段階と、

検索用キーワードとその重みの入力時に前記検索用キーワードに対応するメモリ領域を参照して当該メモリ領域に存する文書の識別情報とその重みとを抽出し、該抽出結果に基づいて前記検索用キーワードに関連する文書の識別情報を関連度順に出力する段階と、を含むことを特徴とする文書検索方法。

【請求項２】　前記検索対象となる文書に含まれるキーワードと該キーワードが関連する１以上の他のキーワードの関連度とを、前記他のキーワードをもとに逆引きできるように予め編集しておき、一のキーワードとその重みが入力されたときに、該入力キーワードが関連するキーワードとその関連度を当該入力キーワードから逆引きして取得し、取得したキーワードとその関連度、及び前記入力キーワードとその重みから、前記各キーワードに対する文書の重みを算出することを特徴とする請求項１記載の文書検索方法。

【請求項３】　検索用キーワードと検索目的に応じて当該検索用キーワードに付与された重み、及び、前記検索用キーワードに関連する１以上の他のキーワードとその関連度から前記検索用キーワードに対する検索対象となる文書の関連度を決定する文書関連度決定部を有する文書検索装置において、

予め前記文書の識別情報、該文書に含まれるキーワードと検索目的に応じて当該キーワードに付与された重み、及び、前記キーワードが関連する他のキーワードとその関連度に基づいて算出された当該文書の重みを当該文書の識別情報と共にキーワード毎に格納した拡張文書エントリ辞書を備え、

前記文書関連度決定部が、前記文書の関連度を決定する際に、前記拡張文書エントリ辞書から前記検索用キーワードについて格納されている文書の識別情報と当該文書の重みとを抽出することを特徴とする文書検索装置。

【請求項４】　前記文書関連度決定部は、前記検索用キーワードとの関連度の高い順に文書の識別情報を出力することを特徴とする請求項３記載の文書検索装置。

【請求項５】　検索対象となる文書に含まれるキーワードと該キーワードが関連する１以上の他のキーワードの関連度とを前記他のキーワードをもとに逆引きできるように予め編集されたシソーラス辞書と、

一のキーワードと検索目的に応じて前記一のキーワードに付与された重みが入力されたときに、該入力キーワードが関連するキーワードとその関連度を前記シソーラス辞書から逆引きして取得し、取得したキーワードとその関連度、及び前記入力キーワードとその重みから、前記各キーワードに対する文書の重みを算出するとともに、算出された文書の重みと当該文書の識別情報とをキーワード毎に対応付けた拡張文書エントリ辞書を生成する辞書生成手段と、を備え、

検索用キーワードとその重みの入力時に、前記生成された拡張文書エントリ辞書から前記検索用キーワードに対応付けられた文書の識別情報とその重みが抽出されるように構成された文書検索装置。

【請求項６】　前記辞書生成手段は、キーワードとその重み、及び、前記キーワードが関連する他のキーワードとその関連度を取得する度に、各キーワードに対する文書の重みを算出し、算出された重みを当該文書の識別情報と共に前記拡張文書エントリ辞書の同一キーワードの設定領域へ連続的に蓄積することを特徴とする請求項４記載の文書検索装置。

【請求項７】　検索対象となる文書の識別情報、当該文書に含まれるキーワードと検索目的に応じて当該キーワードに付与される重み、及び、前記キーワードが関連する他のキーワードとその関連度を取得して各キーワードに対する文書の重みを算出する処理、

算出された重みを当該文書の識別情報と共にキーワード毎に設定された所定のメモリ領域へ蓄積する処理、

検索用キーワードとその重みの入力時に前記検索用キーワードに対応する前記メモリ領域を参照して当該メモリ領域に存する文書の識別情報とその重みとを抽出し、該抽出結果に基づいて前記検索用キーワードに関連する文書の識別情報を関連度順に出力する処理、をコンピュータに実行させるためのプログラムを当該コンピュータが読取可能な形態で記録してなる記録媒体。

【請求項８】　前記プログラムが、さらに、前記検索対象となる文書に含まれるキーワードと該キーワードが関連する１以上の他のキーワードの関連度とを、前記他のキーワードをもとに逆引きできるように編集する処理、一のキーワードとその重みが入力されたときに、該入力キーワードが関連するキーワードとその関連度を当該入力キーワードから逆引きして取得し、取得したキーワードとその関連度、及び前記入力キーワードとその重みから、前記各キーワードに対する文書の重みを算出する処理を、コンピュータに実行させることを特徴とする請求項７記載の記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、例えば大量の文書や文を蓄積した文書データベース、蓄積された文字等の情報を文書作成や発想展開の支援に用いる各種支援システム等に適用される文書検索技術に係り、特に、入力された検索用キーワードの他に、その検索用キーワードに

関連する他のキーワードをも検索語に加えて該当する文書等を検索する文書検索技術に関する。

【０００２】

【従来の技術】検索対象となる文書や文（以下、文書とする）を格納した文書データベース等から所要の文書を検索するコンピュータ（以下、文書検索装置）において、例えば文書に含まれる検索単語から成るキーワードのみならず、検索目的に応じて付与された重みも入力することで、検索精度を高めることが試みられている。

【０００３】図８は、この種の従来の文書検索装置の構成図である。この文書検索装置８は、個々の文書に含まれるキーワードと当該キーワードに関連する１以上の他のキーワードとをその関連度とともに格納したシソーラス辞書８１と、文書識別情報の一例である文書番号とその重みとをキーワード毎に文書エントリとして格納した文書エントリ辞書８２とを備えている。

【０００４】また、上記シソーラス辞書８１と文書エントリ辞書８２とを参照して検索を行うために、シソーラス展開部８３、文書関連度決定部８４、及び文書エントリ検索部８５を備えている。データ量が大きな文書エントリ辞書８２については、ハードディスク等の大容量の補助記憶媒体に記録される。

【０００５】シソーラス展開部８３は、外部から検索用キーワードとその重み（例えば重要度に応じて付加される数値）の入力を受け付けるとともに、シソーラス辞書８１を参照して、検索用キーワードに対応するキーワードを検索し、検索用キーワードと関連キーワードとの関連度に当該検索用キーワードの重みを掛け合わせた関連キーワードの重みを算出する。そして、文書関連度決定部８４に検索用キーワードとその重み、及び関連キーワードとその重みを送る。以後、検索用キーワードとそれに対応する関連キーワードとの組をシソーラス展開されたキーワードと称する。

【０００６】文書関連度決定部８４は、このシソーラス展開されたキーワードを文書エントリ検索部８５に与える。文書エントリ検索部８５は、キーワードと文書エントリ辞書８２内のメモリ領域のアドレスとの対応関係を記述したアドレステーブル８６を保持しており、このアドレステーブル８６を参照して、シソーラス展開されたキーワードに対応するアドレスをキーワード毎に取得する。そして、キーワード毎に取得した文書エントリ辞書８２のアドレスに存する文書エントリを索出するとともに、それを文書関連決定部８４に返信する。文書関連度決定部８４は、文書エントリ検索部８５から受け取った文書エントリ（ここでは文書番号）毎に、文書と検索用キーワードとの関連度を算出し、これを検索結果として出力する。

【０００７】ここで、文書と検索用キーワードとの関連度について説明する。前述のように、文書関連度決定部８４は、文書エントリ検索部８５からシソーラス展開されたキーワード毎に、対応する文書エントリを取得する。これは逆にみれば、文書番号毎に対応する複数のキーワード（シソーラス展開されたキーワード）を取得したことを意味する。そして、文書と検索用キーワードとの関連度は、当該文書番号に対応する一つ以上のキーワードの重みに当該文書番号で表された文書の重みを掛け合わせて得られる数の総和で与えられる。

【０００８】なお、上記関連度の計算、キーワードの重み、文書中のキーワードの重みは、ＴＦ／ＩＤＦ法等のアルゴリズムに基づいて決定される。このＴＦ／ＩＤＦ法についての詳細については、「Intorduction to Modern Information Retrieval」（Gerald Salton他著、MacGraw-Hill Publishing Company）の記載を参考にすることができる。

【０００９】

【発明が解決しようとする課題】上記従来の文書検索装置８には、シソーラス辞書８１に、個々のキーワードについて多くの関連キーワードが格納されている場合、シソーラス展開部８３によって検索されるキーワード数が多くなり、これに伴って、文書エントリ検索部８５が文書エントリ辞書８２において検索するキーワード数が多くなる傾向があった。また、文書エントリ辞書８２が大容量の補助記憶媒体に記録されている場合、文書エントリ検索部８５が文書エントリ辞書８２から文書エントリを検索する際に、適当なシーク時間（アクセスを要求してからアクセス可能になるまでの時間）が必要となる。しかし、一般に大容量の補助記憶媒体のシーク動作は非常に低速であるため、検索速度は、文書エントリ検索部８５に渡されるキーワード数の増加に伴って著しく低下する問題があった。

【００１０】そこで、本発明の課題は、キーワードに関連する他のキーワードの数が増大した場合であっても検索速度の低下を防止することができる文書検索方法を提供することにある。本発明の他の課題は、上記文書検索方法の実施に適した文書検索装置、及びこの文書検索方法を汎用のコンピュータを用いて実現するための記録媒体を提供することにある。

【００１１】

【課題を解決するための手段】上記課題を解決する本発明の文書検索方法は、検索対象となる文書の識別情報、当該文書に含まれるキーワードと検索目的に応じて当該キーワードに付与される重み、及び、前記キーワードが関連する他のキーワードとその関連度を取得して各キーワードに対する文書の重みを算出し、算出された重みを当該文書の識別情報と共にキーワード毎に設定されたメモリ領域へ蓄積する段階と、検索用キーワードとその重みの入力時に前記検索用キーワードに対応するメモリ領域を参照して当該メモリ領域に存する文書の識別情報とその重みとを抽出し、該抽出結果に基づいて前記検索用キーワードに関連する文書の識別情報を関連度順に出力

する段階と、を含むことを特徴としている。

【００１２】上記方法において、好ましくは、検索対象となる文書に含まれるキーワードと該キーワードが関連する１以上の他のキーワードの関連度とを、前記他のキーワードをもとに逆引きできるように予め編集しておき、一のキーワードとその重みが入力されたときに、該入力キーワードが関連するキーワードとその関連度を当該入力キーワードから逆引きして取得し、取得したキーワードとその関連度、及び前記入力キーワードとその重みから、前記各キーワードに対する文書の重みを算出するようにする。

【００１３】上記他の課題を解決するため、本発明は、検索用キーワードと検索目的に応じて当該検索用キーワードに付与された重み、及び、前記検索用キーワードに関連する１以上の他の関連キーワードとその関連度から前記検索用キーワードに対する検索対象となる文書の関連度を決定する文書関連度決定部を有する文書検索装置において、予め前記文書の識別情報、該文書に含まれるキーワードと検索目的に応じて当該キーワードに付与された重み、及び、前記キーワードが関連する他の関連キーワードとその関連度に基づいて算出された当該文書の重みを当該文書の識別情報と共にキーワード毎に格納した拡張文書エントリ辞書を設け、前記文書関連度決定部が、前記文書の関連度を決定する際に、前記拡張文書エントリ辞書から前記検索用キーワードについて格納されている文書の識別情報と当該文書の重みとを抽出し、必要に応じて前記検索用キーワードとの関連度の高い順に文書の識別情報を出力する文書検索装置を提供する。

【００１４】本発明は、また、検索対象となる文書に含まれるキーワードと該キーワードが関連する１以上の他のキーワードの関連度とを前記他のキーワードをもとに逆引きできるように予め編集されたシソーラス辞書と、一のキーワードと検索目的に応じて前記一のキーワードに付与された重みが入力されたときに、該入力キーワードが関連するキーワードとその関連度を前記シソーラス辞書から逆引きして取得し、取得したキーワードとその関連度、及び前記入力キーワードとその重みから、前記各キーワードに対する文書の重みを算出するとともに、算出された文書の重みと当該文書の識別情報とをキーワード毎に対応付けた拡張文書エントリ辞書を生成する辞書生成手段と、を備え、検索用キーワードとその重みの入力時に、前記生成された拡張文書エントリ辞書から前記検索用キーワードに対応付けられた文書の識別情報とその重みが抽出されるように構成された文書検索装置をも提供する。

【００１５】後者の文書検索装置において、前記辞書生成手段は、例えば、キーワードとその重み、及び、前記キーワードが関連する他のキーワードとその関連度を取得する度に、各キーワードに対する文書の重みを算出し、算出された重みを当該文書の識別情報と共に前記拡張文書エントリ辞書の同一キーワードの設定領域へ連続的に蓄積するように構成される。

【００１６】上記他の課題を解決する本発明の記録媒体は、下記の処理をコンピュータに実行させるためのプログラムがコンピュータ読取可能な形態で記録された記録媒体である。

（１）検索対象となる文書の識別情報、当該文書に含まれるキーワードと検索目的に応じて当該キーワードに付与される重み、及び、前記キーワードが関連する他のキーワードとその関連度を取得して各キーワードに対する文書の重みを算出する処理、

（２）算出された重みを当該文書の識別情報と共にキーワード毎に設定された所定のメモリ領域へ蓄積する処理、

（３）検索用キーワードとその重みの入力時に前記検索用キーワードに対応する前記メモリ領域を参照して当該メモリ領域に存する文書の識別情報とその重みとを抽出し、該抽出結果に基づいて前記検索用キーワードに関連する文書の識別情報を関連度順に出力する処理。

【００１７】好ましくは、さらに、下記の処理もコンピュータに実行させるようにする。

（４）前記検索対象となる文書に含まれるキーワードと該キーワードが関連する１以上の他のキーワードの関連度とを、前記他のキーワードをもとに逆引きできるように編集する処理、

（５）一のキーワードとその重みが入力されたときに、該入力キーワードが関連するキーワードとその関連度を当該入力キーワードから逆引きして取得し、取得したキーワードとその関連度、及び前記入力キーワードとその重みから、前記各キーワードに対する文書の重みを算出する処理。

【００１８】

【発明の実施の形態】以下、図面を参照して、本発明の実施形態を詳細に説明する。図１は、本発明の一実施形態に係る文書検索装置のブロック構成図である。この文書検索装置１は、コンピュータによって実現されるもので、コンピュータの内部あるいは外部記憶装置内に設けられる拡張文書エントリ辞書１３、シソーラス辞書１４、逆引きシソーラス辞書１５、及び、そのコンピュータが所定のプログラムを読み込んで実行することにより形成される、文書関連度決定部１１、文書エントリ検索部１２、逆引きシソーラス辞書作成部１６、文書エントリ登録部１７の機能ブロックを備えて構成される。逆引きシソーラス辞書作成部１６は、さらに読み出し処理部１６１、キーワード等登録処理部１６２の機能要素を含み、文書エントリ登録部１７は、検索処理部１７１、文書エントリ登録処理部１７２の機能要素を含んでいる。符号１７３，１８はアドレステーブルである。また、図示しないが、装置、検索結果を利用者等に提示するための出力装置をも備えている。

【００１９】上記プログラムは、通常、上記内部記憶装置あるいは外部記憶装置に格納され、随時読み取られて実行されるようになっているが、コンピュータとは分離可能な記録媒体、例えばＣＤ－ＲＯＭやＦＤ等の可搬性記録媒体、あるいは当該コンピュータ装置と構内ネットワークを通じて接続されたプログラムサーバ等に格納され、使用時に上記内部記憶装置または外部記憶装置にインストールされて随時実行に供されるものであってもよい。

【００２０】拡張文書エントリ辞書１３には、検索対象となる文書の文書番号と重みとが文書エントリとして格納されており、シソーラス辞書１４には、キーワードとそれに関連する１以上の他のキーワード（関連キーワード）がその関連度と共に格納されている。逆引きシソーラス辞書１５は、上記シソーラス辞書１４の逆引き情報を格納するためのものである。

【００２１】文書関連度決定部１１は、外部から検索用キーワードとその重みの入力を受け付け、受け付けた検索用キーワードを文書エントリ検索部１２に送る。文書エントリ検索部１２は、予めキーワードと拡張文書エントリ辞書１３の格納領域のアドレスとの対応関係を記述したアドレステーブル１８を保持しており、文書関連度決定部１１から検索用キーワードを受け取ったとき、このアドレステーブル１８を参照して、検索用キーワードに対応する拡張文書エントリ辞書１３から該当の文書エントリを読み出し、これを文書関連度決定部１１に返信する。文書関連度決定部１１は、文書エントリ検索部１２から返信された文書エントリ（ここでは文書番号）毎に、文書と検索用キーワードとの関連度を算出し、これを検索結果として出力する。

【００２２】ところで、文書検索装置１において上記文書検索を行う場合は、拡張文書エントリ辞書１３への文書エントリの登録が完了していることが必要となる。そのため、文書検索装置１では、シソーラス辞書１４を参照して逆引きシソーラス辞書作成部１６で逆引きシソーラス辞書１５を作成しておく。そして、文書エントリ登録部１７で、入力されたキーワード及びその重みを上記逆引きシソラーラス辞書１５を用いて拡張文書エントリ辞書１３へ登録する。

【００２３】まず、逆引きシソーラス辞書作成部１６における処理を図２及び図３を参照して具体的に説明する。逆引きシソーラス辞書作成部１６は、図２に示すように、読み出し処理部１６１においてシソーラス辞書１４の内容を読み出す。図３（ａ）は、ここで読み出されるシソーラス辞書１４の内容例を示す図表Ａ１であり、キーワード「ｋｗｄ１」に関連する関連キーワードが「ｋｗｄ４」と「ｋｗｄ７」で、「ｋｗｄ１」との関連度がそれぞれ“０．８”と“０．４”であることを示している。なお、図３（ａ）において、キーワード項目はハッシュキーであり、関連キーワード項目はハッシュ結果を示すものである。他の関連キーワードについても同様な対応関係が与えられる。

【００２４】読み出し処理部１６１は、シソーラス辞書１４から読み出した上記内容に基づいて、関連キーワードからキーワードへの逆の対応関係（逆引きの関係）を与える。図３（ｂ）は、このような対応関係の一例を示す図表Ａ２である。つまり、図３（ａ）によれば、シソーラス辞書１４において、キーワード「ｋｗｄ１」には関連度が“０．８”の関連キーワード「ｋｗｄ４」と、関連度が“０．４”の関連キーワード「ｋｗｄ７」とが対応していた。これを逆にみると、関連キーワード「ｋｗｄ４」は関連度“０．８”のキーワード「ｋｗｄ１」が対応し、関連キーワード「ｋｗｄ７」は関連度“０．４”のキーワード「ｋｗｄ１」に対応したものとなっている。図３（ｂ）は、こうした逆の対応関係を示している。

【００２５】なお、図３（ｂ）では、一つの関連キーワードに一つのキーワードを対応させている。しかし、一つの関連キーワード「ｋｗｄ４」に上記キーワード「ｋｗｄ１」のほかに他のキーワード「ｋｗｄ１０」も対応する場合がある。そこで、キーワード等登録処理部１６２は、このような対応関係を整理して、一の関連キーワードに複数のキーワードが対応する場合は、キーワード項目に複数のキーワードとその関連度とを格納する。この図表Ａ３から、関連キーワード「ｋｗｄ４」は、“関連度”０．８”でキーワード「ｋｗｄ１」と対応し、関連度“０．２”でキーワード「ｋｗｄ１０」に対応していることがわかる。

【００２６】次に、逆引きシソーラス辞書１５が作成された後の文書エントリ登録部１７による登録処理を図４及び図５を参照して説明する。文書エントリ登録部１７は、図４に示すように、登録対象となる文書の文書番号と、当該文書についてＴＦ／ＩＤＦ法等により、あるいは他の手法によって求めたキーワード及び重みとの組を検索処理部１７１に入力する。図５（ａ）は、入力されたキーワード及びその重みの例を示す図表Ｂ１１、文書番号の例を示す図表Ｂ１２である。図示の例では、文書番号として“１２０”、この文書番号“１２０”の文書に含まれるキーワード「ｋｗｄ４」，「ｋｗｄ５」の重みが、それぞれ“３”，“２”となっている。

【００２７】検索処理部１７１は、この入力キーワード「ｋｗｄ４」，「ｋｗｄ５」に対応するキーワードを逆引きシソーラス辞書１５から索出し、各キーワード毎の重みを算出する。このとき、各キーワードの重みは、入力キーワードについてはその重み、索出されたキーワードについてはキーワード間の関連度に入力キーワードの重みを掛け合わせたものである。

【００２８】例えば、逆引きシソーラス辞書１５の内容が図３（ｃ）で示されたものであるときの検索処理部１７１の検索処理の結果は、図５（ｂ）の図表Ｂ２のよう

になる。つまり、図３（ｃ）によれば、入力キーワード「ｋｗｄ４」に対応するキーワードは「ｋｗｄ１」，「ｋｗｄ１０」である。そして、各キーワード「ｋｗｄ１」，「ｋｗｄ１０」の関連度がそれぞれ“０．８”，“０．２”であるので、これらに入力キーワード「ｋｗｄ４」の重み“３”を掛け合わせて、“２．４”，“０．６”となる。同様にして、他の入力キーワード「ｋｗｄ５」についても、キーワード「ｋｗｄ５」の重み“２”と、これに関連するキーワード「ｋｗｄ２」，「ｋｗｄ１」の重み“０．８”，“０．２”を算出する。

【００２９】文書エントリ登録部１７２は、キーワードと拡張文書エントリ辞書１３の格納領域のアドレスの対応を示したアドレステーブル１７３を参照して、キーワード毎に、文書番号と上記重みを登録する。図５（ｃ）はアドレステーブル１７３の内容例を示す図表Ｂ３である。

【００３０】このテーブル１７３では、キーワード「ｋｗｄ１」，「ｋｗｄ２」，…に、それぞれ拡張文書エントリ辞書１３の先頭アドレス“１”、“４０００”、…が対応付けられている。つまり、アドレス“１”～”３９９９”がキーワード「ｋｗｄ１」の領域、“４０００”～“”がキーワード「ｋｗｄ２」の領域、…である。文書エントリ登録部１７２は、図５（ｄ）の図表Ｂ４に示されるように、キーワード「ｋｗｄ１」の領域に文書番号“１２０”と重み“２．４”を文書エントリとして登録し、キーワード「ｋｗｄ２」の領域に文書番号“１２０”と重み“０．８”を文書エントリとして登録する。他のキーワード「ｋｗｄ３」，…についても同様にして登録を済ませる。

【００３１】なお、ここでは関連キーワードからキーワードの検索が可能な逆引きシソーラス辞書１５を使用したが、既存のシソーラス辞書１４を、さらに関連キーワードからキーワード検索が可能になるように作成すれば、逆引きシソーラス辞書１５の代わりに両方向からの検索が可能に作成されたシソーラス辞書１４を用いることができる。

【００３２】次に、図６及び図７を参照して、上記文書検索装置１を用いた文書検索方法について説明する。文書検索装置１は、文書関連度決定部１１において検索用キーワードとその重みの入力を受け付け（ステップＳ１０１）、検索用キーワードについては、それを文書エントリ検索部１２に送る。図７（ａ）は入力された検索用キーワードと重みの一例を示す図表Ｃ１である。ここには、３つの検索用キーワード「ｋｗｄ１」，「ｋｗｄ２」，「ｋｗｄ５」と、各検索用キーワードの重み“１”，“２”，“５”が与えられている。この場合、検索されるのは、これらの３つの検索用キーワードとの関連度が総合的に高い文書（その識別情報）である。

【００３３】文書エントリ検索部１２は、アドレステーブル１８を参照して、検索用キーワードに対応する拡張文書エントリ辞書１３のアドレスからすでに登録された文書エントリを読み出す（ステップＳ１０２）。そして、読み出した文書エントリを文書関連度決定部１１に返信する。ここで、アドレステーブル１８の内容は、すでに図５（ｃ）で紹介したアドレステーブル１７３の内容を示す図表Ｂ３と同一とし、拡張文書エントリ辞書１３の内容は、図５（ｄ）に示した図表Ｂ４と同一とする。このとき、図７（ａ）に示された検索用キーワードが文書エントリ検索部１２に送られると、文書エントリ検索部１２は、アドレステーブル１８を参照して、拡張文書エントリ辞書１５におけるキーワード「ｋｗｄ１」の領域から、そこに登録された文書エントリ、すなわち、文書番号“２４”、“１２０”、“１２”、…とそれぞれ文書に与えられた重み“３．４”、“２．４”、“１．２”を読み出す。

【００３４】キーワード「ｋｗｄ２」についても同様に、拡張文書エントリ辞書１５におけるキーワード「ｋｗｄ２」の領域から、そこに登録された文書番号“１００２”、“６４”、“１２０”、…とそれぞれ文書に与えられた重み“２．４”，“１．２”，“０．８”を読み出す。キーワード「ｋｗｄ３」についても同様の処理を行う。図７（ｂ）は、このようにして文書関連度決定部１１が文書エントリ検索部１２を介して取得したキーワードとその重み、そして、検索された文書の文書番号、文書の重みとの対応関係を示した図表Ｃ２である。

【００３５】文書関連度決定部１１は、文書エントリ検索部１２からキーワード毎の文書エントリを受け取った後、その文書エントリで表された文書番号毎に、文書と検索用キーワードとの関連度を算出し（ステップＳ１０３）、関連度の大きなものから順に文書番号を関連度とともに検索結果として出力する（ステップＳ１０４）。ステップＳ１０４での処理を図７（ｂ）の内容を例に挙げて説明する。

【００３６】いま、文書番号“２４”の文書に注目したとき、重みが“１”のキーワード「ｋｗｄ１」における文書番号“２４”の文書の重みは“３．４”、重みが“２”のキーワード「ｋｗｄ２」における文書番号“２４”の文書の重みは“１．３”、そして重みが“５”のキーワード「ｋｗｄ３」における文書番号“２４”の文書の重みは“０．８”である。こうしてキーワード「ｋｗｄ１」からの寄与“３．４”、キーワード「ｋｗｄ２」からの寄与“２．６”、キーワード「ｋｗｄ１」からの寄与“４．０”を合算すると全体で“１０．０”となる。これが文書番号“２４”の文書の検索用キーワードに対する関連度である。他の文書についても同様にして、文書と検索用キーワードとの関連度が算出される。

【００３７】図７（ｃ）は、文書関連度決定部１１の検索結果例を示す図表Ｃ３である。ここでは、文書番号“２４”の文書が最も大きな関連度“１０．０”をも

ち、文書番号“１２”の文書が次に大きな関連度“８．９”をもつことを示している。以上が文書検索装置１の検索処理についての説明である。

【００３８】拡張文書エントリ辞書１３がハードディスク等の補助記憶媒体に格納された場合、文書エントリ検索部１２がこの補助記憶媒体にアクセスする際に、適当なシーク時間を必要とする。そして、検索所要時間は、おおよそこのシーク時間の合計で決まる。この事実に注目し、本実施形態による文書検索装置１の方が従来技術による文書検索装置８より検索速度において向上していることを説明する。

【００３９】なお、説明を簡単にするため、シソーラス辞書８１，１４には、それぞれキーワードとして唯一のキーワード「ｋｗｄ」と、それに関連する他のキーワードとしてｎ個のキーワード「ｋｗｄ１」～「ｋｗｄｎ」があるものとする。また、文書エントリ辞書８１にはキーワード「ｋｗｄ」、「ｋｗｄ１」～「ｋｗｄｎ」のそれぞれについて文書エントリがｍ個登録され、拡張文書エントリ辞書１３にはキーワード「ｋｗｄ」に対して文書エントリが（ｎ＋１）・ｍ個登録されているものとする。さらに文書エントリ辞書８１と拡張文書エントリ辞書１３に対する文書エントリ検索部８５，１２のシーク時間をいずれもｔ１とし、文書エントリ辞書８１と拡張文書エントリ辞書１３から一つの文書エントリを読み出すのに必要な時間をいずれもｔ２とする。

【００４０】このとき、キーワード「ｋｗｄ」を検索用キーワードとしたとき、従来の文書検索装置８において検索に要する時間を算出する。この場合、シソーラス展開部８３から文書関連度決定部８４に渡されるキーワード数（＝キーワード「ｋｗｄ」についてシソーラス展開したときのキーワードの数）は、ｎ＋１個である。このｎ＋１個のそれぞれについて、文書エントリ辞書８１からｍ個の文書番号を読み出すことになるが、一つのキーワードについて必要なシーク時間はｔ１＋ｍｔ２なので、シーク時間の合計は、（ｎ＋１）・（ｔ１＋ｍｔ２）（＝Ｔ１）となる。これが文書検索装置７の検索所要時間である。

【００４１】一方、本実施形態の文書検索装置１における検索所要時間を算出する。この場合、文書関連度決定部１１から文書エントリ検索部１２に渡されるキーワード「ｋｗｄ」は一個である。従って、このキーワード「ｋｗｄ」について、シーク時間の合計はｔ１＋（ｎ＋１）・ｍｔ２（＝Ｔ２）となる。こうしてＴ１－Ｔ２＝ｎｔ１＞０が導かれる。

【００４２】なお、本実施形態では、シソーラス辞書８１とシソーラス辞書１４にキーワードとして唯一のキーワード「ｋｗｄ」、それに関連する関連キーワードとして「ｋｗｄ１」～「ｋｗｄｎ」があるような簡単な場合を説明したが、より複雑な場合にも同様な計算を行うことで、本実施形態による文書検索装置１が文書検索装置８より検索速度において向上していることを確認することができる。

【００４３】また、拡張文書エントリ辞書１３は、文書エントリ辞書８１と比べて登録された文書エントリが増大するが、各文書エントリをキーワード毎の領域に連続して蓄積（登録）しているので、文書エントリの取得時に補助記憶媒体から連続的にこれを読み出すことができ、実質的な速度の低下は防止されている。また、拡張文書エントリ辞書１３から一つの文書エントリを取得する際のシーク動作は、文書エントリ辞書８１に対するシーク時間と等しいから、シーク動作が少ない分だけ、処理速度を短縮させることができる。

【００４４】また、本実施形態では、文書エントリ検索部１２及び文書エントリ登録処理部１７２に、それぞれ図５（ｃ）の図表Ｂ３に示した内容のアドレステーブル１８，１７３を備えた例を示したが、このアドレステーブル１８，１７３は、文書エントリ検索部１２や文書エントリ処理部１７２ではなく、拡張文書エントリ辞書１３に共通に備えておくこともできる。あるいはアドレステーブル１８，１７３自体を省略することもできる。

【００４５】さらに、本発明は、文の検索にも応用することができる。このときには、拡張文書エントリ辞書１５に格納する際に、文書番号の代わりに文を識別するための番号を使用すればよい。

【００４６】

【発明の効果】以上の説明から明らかなように、本発明によれば、シソーラス展開したキーワード数が増大した場合にも、従来のように改めてシソーラス展開する必要がないので、文書の検索所要速度を向上させることができる効果がある。

【図面の簡単な説明】

【図１】本発明の一実施形態による文書検索装置のブロック構成図。

【図２】逆引きシソーラス辞書作成部の詳細構成図。

【図３】（ａ）～（ｃ）は逆引きシソーラス辞書の作成過程での具体例を示す図表。

【図４】文書エントリ登録部の詳細構成図。

【図５】（ａ）～（ｄ）は文書エントリ登録処理過程での具体例を示す図表。

【図６】文書検索装置を用いた文書検索方法の手順説明図。

【図７】（ａ）～（ｃ）は本実施形態による文書検索の際の具体例を示す図表。

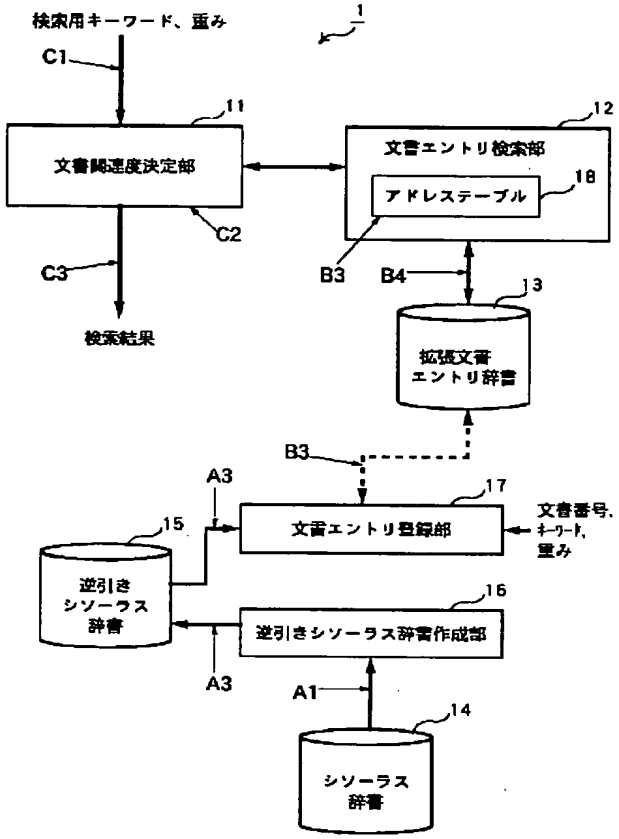
【図８】従来の文書検索装置のブロック構成図。

【符号の説明】

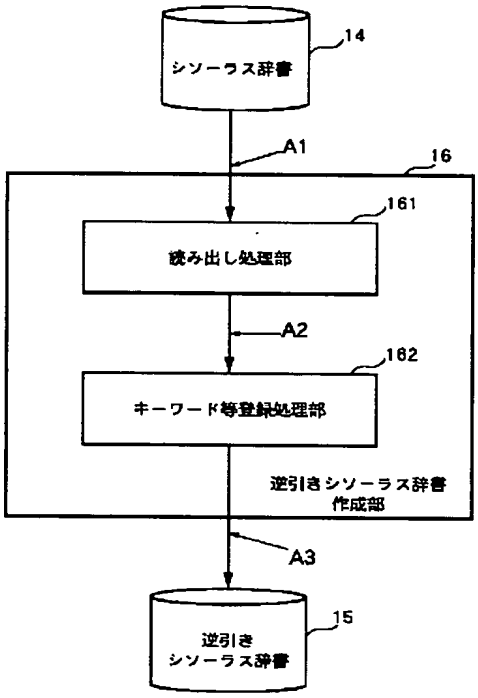
１、８　文書検索装置
１１、８４　文書関連度決定部
１２　文書エントリ検索部
１３　拡張文書エントリ辞書
１４、８１　シソーラス辞書

１５　逆引きシソーラス辞書
１６　逆引きシソーラス辞書作成部
１７　文書エントリ登録部
８２　文書エントリ辞書
８３　シソーラス展開部
８５　文書エントリ検索部
１６１　読み出し処理部
１６２　キーワード等登録処理部
１７１　検索処理部
１７２　文書エントリ登録処理部
１８、１７３　アドレステーブル

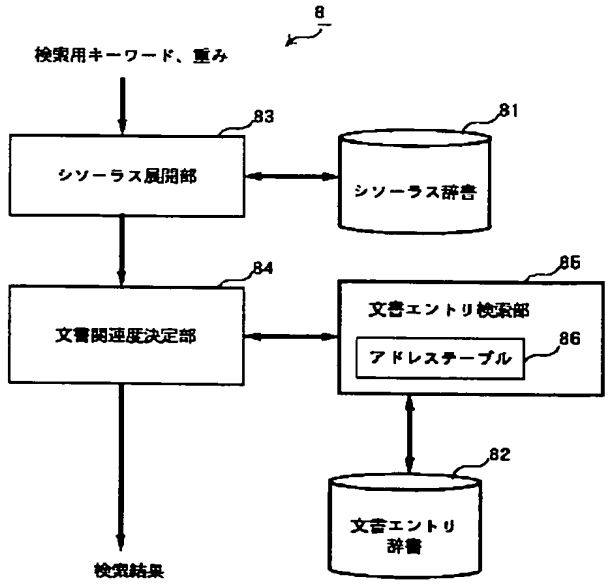
【図１】



【図２】



【図８】



【図３】

(a)

キーワード	関連キーワード（関連度）
kwd1	kwd4 (0.8),kwd7 (0.4)
kwd2	kwd5 (0.4),kwd8 (0.3)
kwd3	kwd6 (0.6),kwd9 (0.5)
・	・
・	・
・	・
kwd10	kwd4 (0.2)
kwd11	kwd5 (0.1)
kwd12	kwd6 (0.3)
・	・
・	・

A1

(b)

関連キーワード（関連度）	キーワード
kwd4 (0.8)	kwd1
kwd5 (0.4)	kwd2
kwd6 (0.6)	kwd3
kwd7 (0.4)	kwd1
kwd8 (0.3)	kwd2
kwd9 (0.5)	kwd3
・	・
・	・
・	・

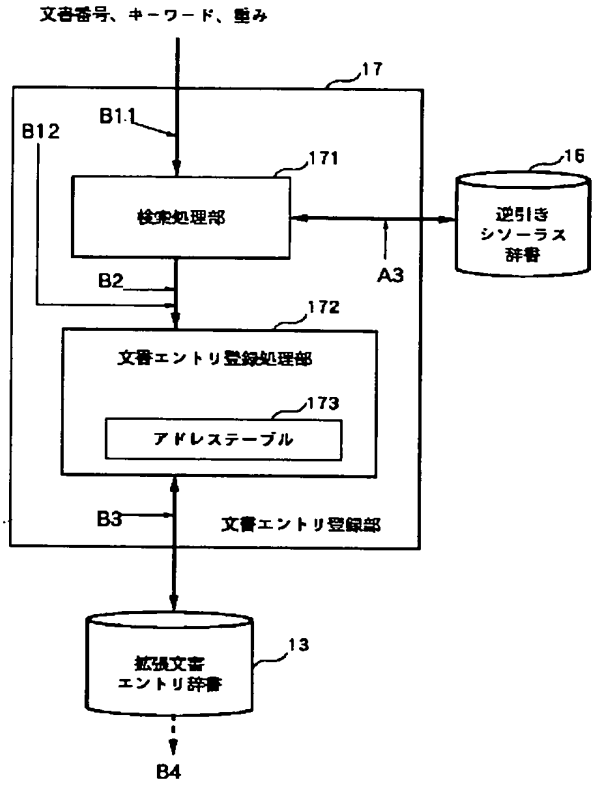
A2

(c)

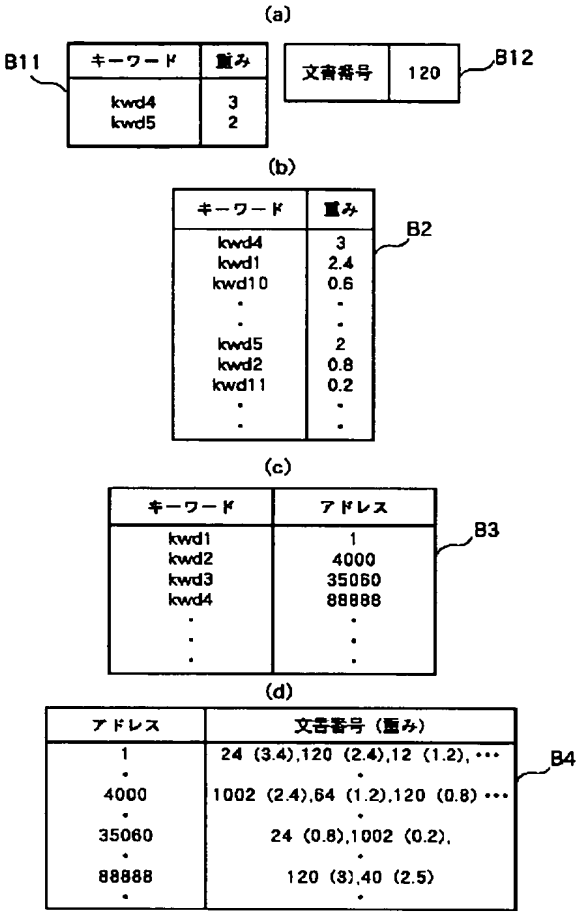
関連キーワード	キーワード（関連度）
kwd4	kwd1 (0.8),kwd10 (0.2)
kwd5	kwd2 (0.4),kwd11 (0.1)
kwd6	kwd3 (0.6),kwd12 (0.3)
・	・
・	・
・	・

A3

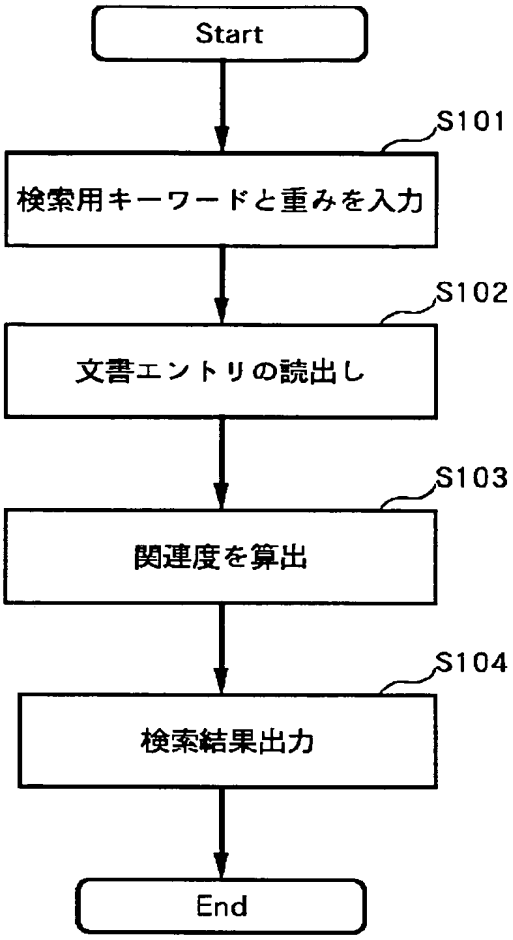
【図４】



【図５】


(a)
B11
キーワード
重み
kwd4
kwd5
3
2
文書番号
120
B12
(b)
キーワード
重み
B2
kwd4
kwd1
kwd10
kwd5
kwd2
kwd11
3
2.4
0.6
2
0.8
0.2
(c)
キーワード
アドレス
B3
kwd1
kwd2
kwd3
kwd4
1
4000
35060
88888
(d)
アドレス
文書番号（重み）
B4
1
4000
35060
88888
24 (3.4),120 (2.4),12 (1.2),・・・
1002 (2.4),64 (1.2),120 (0.8)・・・
24 (0.8),1002 (0.2),
120 (3),40 (2.5)

【図６】


Start
S101
検索用キーワードと重みを入力
S102
文書エントリの読出し
S103
関連度を算出
S104
検索結果出力
End

【図７】

(a)

キーワード	重み
kwd1	1
kwd2	2
kwd3	5

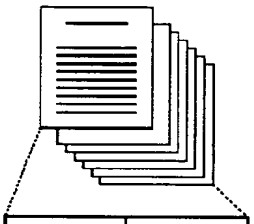
C1

(b)

C2

キーワード	重み	文書番号及び重み
kwd1	1	24 (3.4),12 (1.2), •••
kwd2	2	1002 (2.4),64 (1.2),24 (1.3) •••
kwd3	5	24 (0.8),1002 (0.2), •••

(c)



関連度	文書番号
10.0	24
8.9	12
7.5	1002
6.5	64
・	・
・	・
・	・

C3